



## OPEN ACCESS

## EDITED BY

Catherine Pelachaud,  
UMR7222 Institut des Systèmes  
Intelligents et Robotiques (ISIR), France

## REVIEWED BY

Pierre Raimbaud,  
École Nationale d'Ingénieurs de Saint-  
Etienne, France  
Grégoire Richard,  
University of Technology Compiègne,  
France

## \*CORRESPONDENCE

Tiffany D. Do,  
✉ tiffany.do@ucf.edu

RECEIVED 27 June 2023

ACCEPTED 09 November 2023

PUBLISHED 23 November 2023

## CITATION

Do TD, Zelenty S, Gonzalez-Franco M  
and McMahan RP (2023), VALID: a  
perceptually validated Virtual Avatar  
Library for Inclusion and Diversity.  
*Front. Virtual Real.* 4:1248915.  
doi: 10.3389/frvir.2023.1248915

## COPYRIGHT

© 2023 Do, Zelenty, Gonzalez-Franco  
and McMahan. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original author(s)  
and the copyright owner(s) are credited  
and that the original publication in this  
journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# VALID: a perceptually validated Virtual Avatar Library for Inclusion and Diversity

Tiffany D. Do<sup>1\*</sup>, Steve Zelenty<sup>1</sup>, Mar Gonzalez-Franco<sup>2</sup> and  
Ryan P. McMahan<sup>1</sup>

<sup>1</sup>Department of Computer Science, University of Central Florida, Orlando, FL, United States, <sup>2</sup>Google, Seattle, WA, United States

As consumer adoption of immersive technologies grows, virtual avatars will play a prominent role in the future of social computing. However, as people begin to interact more frequently through virtual avatars, it is important to ensure that the research community has validated tools to evaluate the effects and consequences of such technologies. We present the first iteration of a new, freely available 3D avatar library called the *Virtual Avatar Library for Inclusion and Diversity (VALID)*, which includes 210 fully rigged avatars with a focus on advancing racial diversity and inclusion. We also provide a detailed process for creating, iterating, and validating avatars of diversity. Through a large online study ( $n = 132$ ) with participants from 33 countries, we provide statistically validated labels for each avatar's perceived race and gender. Through our validation study, we also advance knowledge pertaining to the perception of an avatar's race. In particular, we found that avatars of some races were more accurately identified by participants of the same race.

## KEYWORDS

virtual avatars, race, perception, diversity, embodiment

## 1 Introduction

The study of human behavior in virtual worlds is becoming increasingly important as more and more people spend time in these digital spaces. In particular, understanding how behaviors in virtual environments compare to behaviors in real life can provide insight into the ways in which humans interact with technology and each other. Furthermore, it can also help inform the design of virtual environments, making them more realistic and socially acceptable.

One area of significant interest is virtual avatars, which are 3D representations of virtual humans often used in virtual worlds and simulations (Pelachaud, 2009). Virtual avatars have a prominent role in social computing and immersive environments (Yee and Bailenson, 2007; Fox et al., 2015; Heidicker et al., 2017). For example, previous studies have utilized virtual avatars to express emotions (Butler et al., 2017), recreate social psychology phenomena (Gonzalez-Franco et al., 2018), and influence feelings of presence or embodiment (Nowak and Biocca, 2003; Peck and Gonzalez-Franco, 2021). Moreover, virtual avatars are expected to play evermore important roles in the future. For instance, virtual avatars are at the center of every social virtual reality (VR) and augmented reality (AR) interaction, as they are key to representing remote participants (Piumsomboon et al., 2018) and facilitating collaboration (Sra and Schmandt, 2015).

Open virtual avatar libraries facilitate research by providing freely available resources to the community. For instance, Microsoft's Rocketbox avatar library (Gonzalez-Franco et al., 2020b) has been extensively used in multiple studies (e.g., Jeong et al., 2020; Cheng et al., 2022; Huang et al., 2022), with over 400 stars and 100 forks on GitHub since its release in 2020. However, the representation of races in currently available avatar libraries are limited and have not been validated through user perception studies. As examples, over 40% of the Rocketbox avatars are white males, and MakeHuman<sup>1</sup>, another popular resource for 3D avatars, only includes options for African, Asian, and Caucasian avatars. Due to such limitations, many researchers are unable to properly investigate the effects of diversity for virtual avatars. Furthermore, existing libraries not only have limited diversity but are also biased in their lack of inclusivity with regard to professions for minority avatars. For instance, Rocketbox lacks Asian avatars dressed for medical scenarios.

The limited representation of currently available resources may have an inherently detrimental impact by introducing biases into research investigations from the outset, such as restricting the development of scenarios with Asian medical professionals. Furthermore, as immersive technologies grow in popularity worldwide, Taylor et al. (2020) have called for improved racial/ethnic representations in VR to mitigate potential negative effects of racial bias. This can be especially important since virtual avatar race<sup>2</sup> can have a strong influence on human behavior. For example, research has found that embodying a virtual avatar of a different race can affect implicit racial bias, as demonstrated by several prior studies (Groom et al., 2009; Peck et al., 2013; Maister et al., 2015; Salmanowitz, 2018). As such, we specifically focus on providing more racially inclusive avatar resources in this first iteration of the library, with further plans to expand to more inclusive representations.

The goal of this work was dual purpose. First, we sought to develop an open library of validated virtual avatars, which we refer to as the *Virtual Avatar Library for Inclusion and Diversity (VALID)*. This library has been designed to inclusively represent a range of races across various professions. Secondly, we wanted to better understand how humans perceive avatars, particular with regard to race.

In this first iteration of VALID, we followed the recommendations of the 2015 National Content Test Race and Ethnicity Analysis Report from the U.S. Office of Management and Budget (Matthews et al., 2017) to ensure that a diverse range of people was represented in our library. For example, this report found that members of Middle Eastern and North African (MENA) communities use the MENA category when it is available, but have trouble identifying their race when it is not available. Similarly, it also found that Hispanics can better identify their ethnicity by combining the conventional race and Hispanic origin questions into one. Therefore, to overcome those traditional problems on racial and ethnic categorization, VALID includes seven races as

recommended by this report (Matthews et al., 2017): American Indian or Native Alaskan (AIAN)<sup>3</sup>, Asian, Black or African American (Black), Hispanic, Latino, or Spanish (Hispanic), Middle Eastern or North African (MENA), Native Hawaiian or Pacific Islander (NHPI), and White. While these categories may diverge from the latest 2020 census classifications, it is important to note that, as of the time of writing, these categories are in the finalization process under the oversight of the Federal Interagency Technical Working Group on Race and Ethnicity Standards. These forthcoming revisions are anticipated to be incorporated into future updates<sup>4</sup>.

VALID includes 210 fully rigged virtual avatars designed to advance diversity and inclusion. We iteratively created 42 base avatars (7 target races × 2 genders × 3 individuals) using a process that combined data-driven average facial features with extensive collaboration with representative stakeholders from each racial group (see Figure 1 for all base avatars). To address the longstanding issue of the lack of diversity in virtual designers and to empower diverse voices (Boberg et al., 2008), we adopted a participatory design method. This approach involved actively involving individuals from diverse backgrounds ( $n = 22$ ), particularly different racial and ethnic identities, in the design process. By including these individuals as active participants, we aimed to ensure that their perspectives, experiences, and needs were considered and incorporated into the design of the avatars.

Once the avatars were created, we sought to evaluate their perception on a global scale. We then conducted a large online study ( $n = 132$ ) with participants from 33 countries, self-identifying as one of the seven represented races, to determine whether the race and gender of each avatar are recognizable, and therefore validated. We found that all Asian, Black, and White avatars were universally identified as their modeled race by all participants, while our AIAN, Hispanic, and MENA avatars were typically only identified by participants of the same race, indicating that participant race can bias perceptions of a virtual avatar's race. We have since modeled the 42 base avatars in five different outfits (casual, business, medical, military, and utility), yielding a total of 210 fully rigged avatars.

To foster diversity and inclusion in virtual avatar research, we are making all of the avatars in our library freely available to the community as open source models. In addition to the avatars, we are also providing statistically validated labels for the race and gender of all 42 base avatars. Our models are available in FBX format, are compatible with previous libraries like Rocketbox (Gonzalez-Franco et al., 2020b), and can be easily integrated into most game engines such as Unity and Unreal. Additionally, the avatars come equipped with facial blend shapes to enable researchers and developers to easily create dynamic facial expressions and lip-sync animations. All avatars, labels, and metadata can be found at our GitHub repository: <https://github.com/xrtrlab/Validated-Avatar-Library-for-Inclusion-and-Diversity—VALID>.

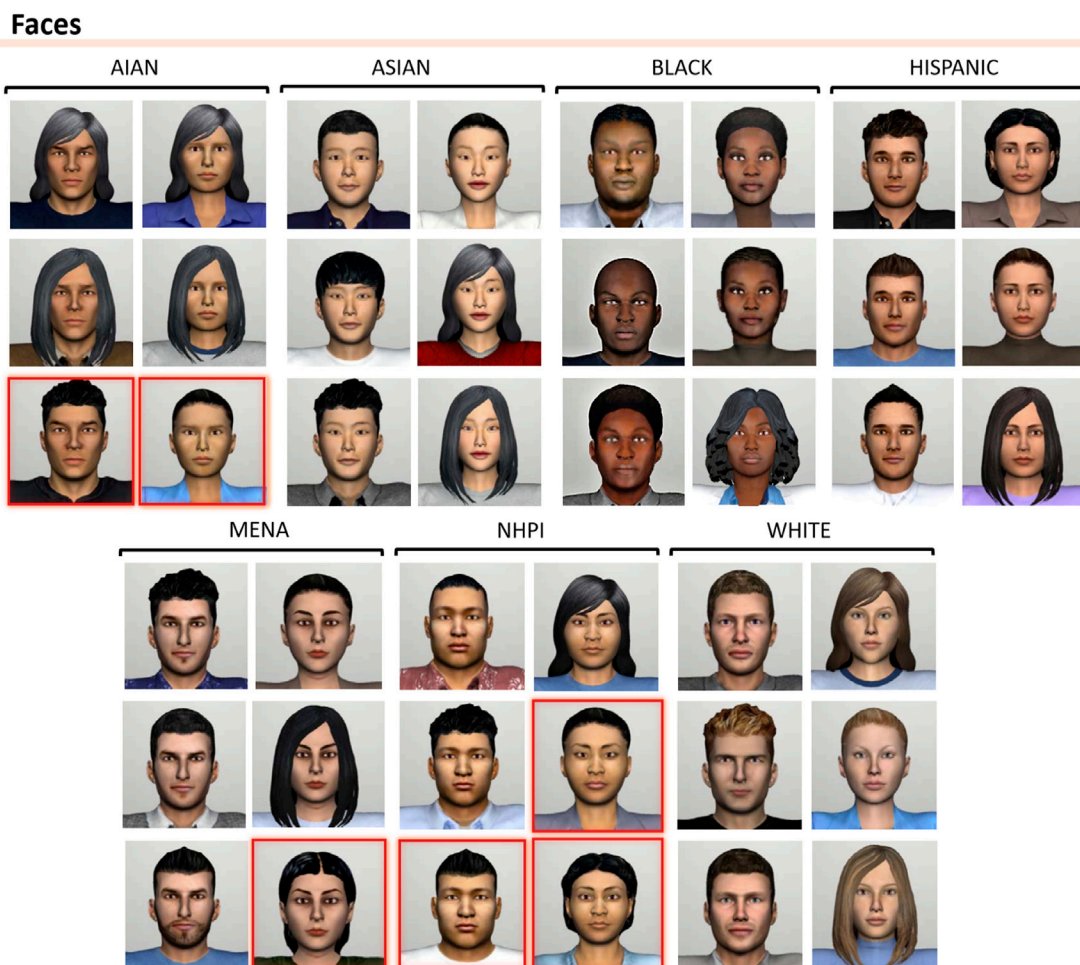
This paper makes three primary contributions:

1 <http://www.makehumancommunity.org/>

2 We use the term race to refer to a "visually distinct social group with a common ethnicity" (Rhodes et al., 2010).

3 We use racial abbreviations as defined in the U.S. Census Bureau report.

4 <https://spd15revision.gov/content/spd15revision/en/news/2023-05-18-release.html>



**FIGURE 1**  
Headshots of all 42 base avatars in the "Casual" outfit. Avatars outlined in red were not validated for their intended race.

- 1) We provide 210 openly available, fully rigged, and perceptually validated avatars for the research community, with a focus on advancing diversity and inclusion.
- 2) Our diversity-represented user study sheds new light on the ways in which people’s own racial identity can affect their perceptions of a virtual avatar’s race. In our repository, we also include the agreement rates of all avatars, disaggregated by every participant race, which offers valuable insights into how individuals from different racial backgrounds perceive our avatars.
- 3) We describe a comprehensive process for creating, iterating, and validating a library of diverse virtual avatars. Our approach involved close collaboration with stakeholders and a commitment to transparency and rigor. This could serve as a model for other researchers seeking to create more inclusive and representative virtual experiences.

## 2 Related work

In this section, we describe how virtual avatars are used within current research in order to highlight the need for diverse avatars.

We conclude the section with a discussion on currently available resources used for virtual avatars and virtual agents.

### 2.1 Effect of avatar race

Virtual avatars are widely used in research simulations such as training, education, and social psychology. The race of a virtual avatar is a crucial factor that can affect the outcomes of these studies. For example, research has shown that underrepresented students often prefer virtual instructors who share their ethnicity (Baylor, 2003; Moreno and Flowerday, 2006). Similarly, studies have suggested that designing a virtual teacher of the same race as inner-city youth can have a positive influence on them (Baylor, 2009), while a culturally relevant virtual instructor, such as an African-American instructor for African-American children, can improve academic achievement (Gilbert et al., 2008).

The design of virtual avatars is especially important for minority or marginalized participants. Kim and Lim (2013) reported that minority students who feel unsupported in traditional classrooms develop more positive attitudes towards avatar-based learning. In addition, children with autism spectrum disorder treat virtual

avatars as real social partners (Ali et al., 2020; Li et al., 2021). Therefore, to better meet the needs of all individuals participating in such studies, it is important for researchers to have access to diverse avatars that participants can comfortably interact with. Diversity in virtual avatars is important not only for improving representation, but also for enhancing the effectiveness of simulations. Halan et al. (2015) found that medical students who trained with virtual patients of a particular race demonstrated increased empathy towards real patients of that race. Similarly, Bickmore et al. (2021) showed that interacting with a minority virtual avatar reduced racial biases in job hiring simulations.

These findings highlight the importance of diverse and inclusive virtual avatars in research simulations and emphasize the need for more comprehensive representation of different races. Access to a wide range of validated avatars through VALID will help to create more inclusive and representative simulations, and enable researchers to investigate the impact of avatar race or gender on participants' experiences. This will help improve the inclusivity of simulations and contribute towards addressing issues of bias.

## 2.2 Implicit racial bias and virtual avatars

Avatars are becoming increasingly important in immersive applications, particularly in the realm of VR, where they are becoming ubiquitous (Dewez et al., 2021). Research suggests that the degree of similarity between a user's real body and their virtual avatar can influence embodiment, presence, and cognition (Jo et al., 2017; Jung et al., 2018; McIntosh et al., 2020; Ogawa et al., 2020). Maister et al. (2015) proposed that non-matching self-avatars can affect a user's self-association and influence social cognition. Moreover, studies have demonstrated that embodying a darker-skinned avatar in front of a virtual mirror can reduce implicit racial biases (Groom et al., 2009; Peck et al., 2013; Maister et al., 2015; Salmanowitz, 2018), which are unconscious biases that can lead to discriminatory behavior (Chapman et al., 2013). For instance, Salmanowitz (2018) found that a VR participant's implicit racial bias affects their willingness to convict a darker-skinned suspect based on inconclusive evidence. Similarly, Peck et al. (2021) found that each participant's implicit racial bias was related to their nuanced head and hand motions in a firearm simulation. These foundational studies provide compelling evidence that embodying an avatar of a different race can affect implicit biases and further emphasize the need for diverse avatar resources.

Our study examines how participants perceive the race of diverse virtual avatars. While some studies have explored how a virtual avatar's race affects user interactions (e.g., Baylor, 2003; Gamberini et al., 2015; Zipp and Craig, 2019), little research has been conducted on how individuals *actively perceive* the race of virtual avatars. Setoh et al. (2019) note that racial identification can predict implicit racial bias, making it crucial to understand how people perceive the race of virtual avatars to further investigate these effects.

## 2.3 Own-race bias

Own-race bias, also known as the "other-race effect," refers to the phenomenon in which individuals process the faces of their own

race differently from those of other races (Meissner and Brigham, 2001; Ge et al., 2009; Rhodes et al., 2009; Civile and McLaren, 2022). Studies have suggested that this bias can influence the way individuals categorize race. For example, MacLin and Malpass (2001) found that Hispanic participants were more likely to categorize Hispanic faces as fitting their racial category than Black faces, and Blascovich et al. (1997) observed that participants who strongly identify with their in-group are more accurate in identifying in-group members.

Although own-race bias has not yet been studied in the context of 3D virtual avatars, Saneyoshi et al. (2022) recently discovered that it extends to the uncanny valley effect (Mori et al., 2012) for 2D computer-generated faces. Specifically, they found that Asian and European participants rated distorted faces of their own race as more unpleasant than those of other races. Building on this research, we extended the study of own-race bias to 3D virtual avatars and focused on race categorization rather than perceived pleasantness. Our study included avatars and participants from seven different races, providing insights into how a diverse user population may interact within equally diverse virtual worlds.

## 2.4 Virtual avatar resources

There are numerous resources for creating virtual avatars. Artists can use 3D modeling tools, such as Autodesk 3ds Max<sup>5</sup>, Autodesk Maya<sup>6</sup>, Blender<sup>7</sup>, or ZBrush<sup>8</sup> to manually model, texture, and rig virtual avatars. However, such work requires expertise in 3D modeling and character design, and is often a tedious process (Gonzalez-Franco et al., 2020b). On the other hand, parametric models, including freely available tools like MakeHuman<sup>9</sup> and Autodesk Character Generator<sup>10</sup>, as well as commercially available ones such as Daz3D<sup>11</sup>, Poser<sup>12</sup>, and Reallusion Character Creator<sup>13</sup>, enable users to generate virtual avatars from predefined parameters, thereby significantly expediting the avatar generation process. Nonetheless, using these tools still requires learning a new program and time to customize each model, despite the absence of the artistic expertise needed for manual tools.

Another alternative to traditional modeling is to use scanning technologies, which can capture 3D models of real people. For instance, Shapiro et al. (2014) and Waltemate et al. (2018) used 3D cameras and photogrammetry, respectively, to capture 3D models of their users. Singular Inversions FaceGen Modeller<sup>14</sup> has also been employed to generate 3D faces from user photos

5 <https://www.autodesk.com/products/3ds-max/overview>

6 <https://www.autodesk.com/products/maya/overview>

7 <https://www.blender.org>

8 <https://www.maxon.net/en/zbrush>

9 <http://www.makehumancommunity.org/>

10 <https://charactergenerator.autodesk.com/>

11 <https://www.daz3d.com>

12 <https://www.posersoftware.com>

13 <https://www.reallusion.com/character-creator/>

14 <https://facegen.com/modeller.htm>



and then apply them to a general 3D avatar body (Blom et al., 2014; Gonzalez-Franco et al., 2016). However, scanning approaches require the ability to physically scan the user, limiting their use for certain applications, particularly remote ones.

Most closely related to our goal of providing a free and open library of ready-to-use avatars is the Microsoft Rocketbox library (Gonzalez-Franco et al., 2020b) and its accompanying HeadBox (Volonte et al., 2022) and MoveBox (Gonzalez-Franco et al., 2020a) toolkits. Rocketbox provides a free set of 111 fully rigged adult avatars of various races and outfits. However, it falls short in terms of representation by not including any avatars of AIAN or NHPI descent. Additionally, the library offers only a limited number of Asian, Hispanic, and MENA avatars, excluding minority representations for some professions (e.g., Rocketbox does not include any Asian medical avatars). Furthermore, none of the available avatar libraries have been validated by user perception studies to ensure their efficacy and inclusivity. Therefore, our VALID project aims to fill this gap by providing a free and validated library of diverse avatars.

### 3 Avatar creation procedure

This section outlines our iterative process for developing the VALID library, which includes 42 base avatars. We began by using data-driven averaged facial features to create our initial models. We then conducted interviews with representative volunteers to iteratively refine and modify the avatars based on their feedback.

#### 3.1 Initial modeling

To ensure a broad diversity of people were represented in our library, we initially created 42 base avatars (7 target races  $\times$  2 genders  $\times$  3 individuals) modeled after the seven racial groups recommended by the 2015 National Content Test Race and Ethnicity Analysis Report (Matthews et al., 2017): AIAN, Asian, Black, Hispanic, MENA, NHPI, and White. We created 3 male and 3 female individuals for each race, resulting in a total of 6 individuals per race.

Preliminary models were based on averaged facial features of multiple photos selected from the 10 k US Adult Faces Database (Bainbridge et al., 2013) and stock photos from Google for races missing from the database (e.g., AIAN, MENA, and NHPI). These photos were used as input to a face-averaging algorithm (DeBruine, 2018), which extracted average facial features for each race and gender pair. Using these averages as a reference, a 3D artist recreated the average faces for each race and gender pair using Autodesk Character Generator (due to its generous licensing and right to freely edit and distribute generated models<sup>15</sup>) and Blender to make modifications not supported by Autodesk Character Generator (see Figure 2).

#### 3.2 Iterative improvements through representative interviews

After the preliminary avatars were created based on the facial averages, we worked closely with 2–4 volunteers of each represented race (see Table 1) to adjust the avatars through a series of Zoom meetings. This process ensured that all avatars were respectful and reduced the likelihood of harmful or stereotypical representations. Volunteers self identified their race and were recruited from university cultural clubs (e.g., Asian Student Association, Latinx Student Association), community organizations (e.g., Pacific Islanders Center), and email lists.

We iteratively asked these volunteers for feedback on all avatars representing their race, showing them the model from three perspectives (see Figure 3). Volunteers were specifically asked to identify accurate features and suggest changes to be made. Each Zoom meeting lasted around 15–30 min on average and concluded once all feedback was recorded. Once the changes were completed based on the feedback, we presented the updated avatars to the volunteers. This process was repeated until they approved the appearance of the avatars. Meetings were held between 1 and 3 times, depending on the amount of changes requested by representatives. For example, volunteers requested changes to facial features, such as:

- “Many Native women I know have a softer face and jawline [than this avatar].” –AIAN Volunteer 3
- “The nose bridge is too high and looks too European. Asians mostly have low nose bridges.” –Asian Volunteer 2
- “Middle Eastern women usually have wider, almond-shaped eyes.” –MENA Volunteer 2
- “The nose [for this avatar] should be a little bit thicker, less pointy, and more round.” –NHPI Volunteer 1

Additionally, we modified hairstyles according to feedback:

- “[These hairstyles] look straighter and more Eurocentric. So I would choose [these facial features] and then do a natural [hair] texture.” –Black Volunteer 1
- “Usually the men have curly hair or their hair is cut short on the sides with the top showing.” –NHPI Volunteer 1

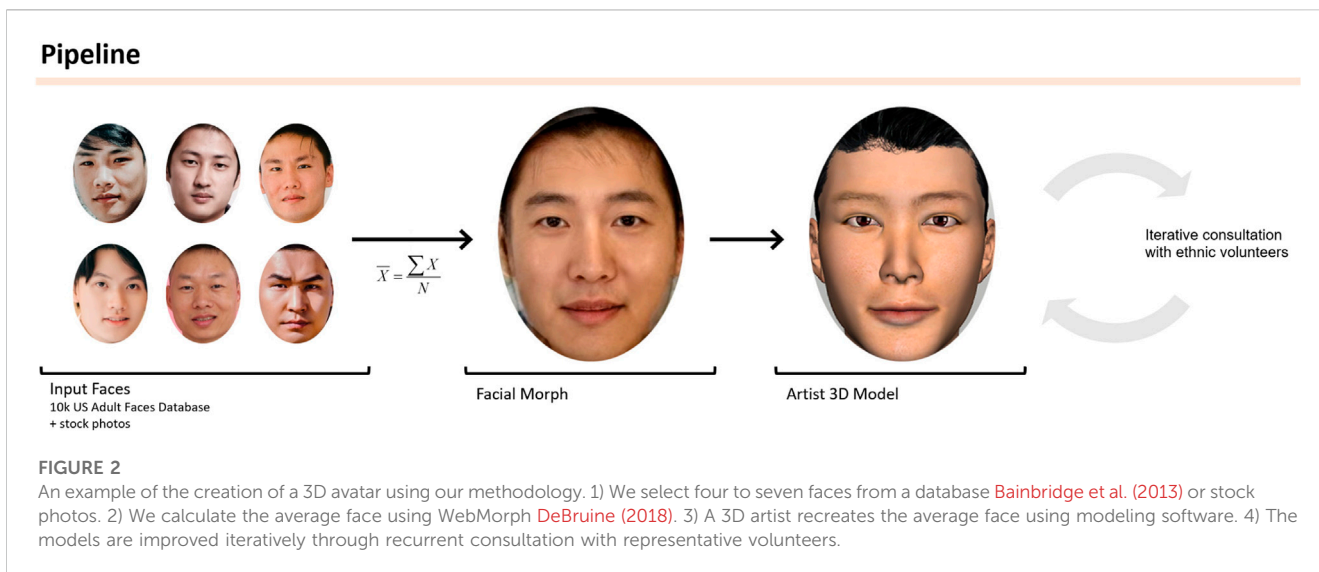
Once the avatars were approved by their corresponding volunteer representatives, we conducted an online study to validate the race and gender of each avatar based on user perceptions.

### 4 Avatar validation study

We conducted an online, worldwide user study to determine whether the target race and gender of each avatar is recognizable and, therefore, validated. Participants were recruited from the online Prolific marketplace<sup>16</sup>, which is

<sup>15</sup> <https://knowledge.autodesk.com/support/character-generator/learn-explore/caas/sfdcarticles/sfdcarticles/Character-Generator-Legal-Restrictions-and-Allowances-when-using-Character-Generator.html>

<sup>16</sup> <https://www.prolific.co>



**TABLE 1 Breakdown of our volunteer representatives by race, gender (male, female, or non-binary), and country.**

Race	Gender	Country
AIAN	2M, 1F	United States (Native American) (3)
Asian	2M, 2F	China (1), United States (Chinese-American) (1), United States (Vietnamese-American) (2)
Black	1M, 2F	United States (African-American) (3)
Hispanic	2M, 1F	Mexico (1), United States (Mexican-American) (2)
MENA	1M, 2F	Iran (2), Saudi Arabia (1)
NHPI	1M, 2F	Samoa (2), United States (Native Hawaiian) (1)
White	2M, 1F	United States (3)

similar to Amazon Mechanical Turk. Prior research shows that Prolific has a pool of more diverse and honest participants (Peer et al., 2017) and has more transparency than Mechanical Turk (Palan and Schitter, 2018). Since diversity was a core theme of our research, we chose Prolific to ensure that our participants would be diverse.

### 4.1 Procedure

The following procedure was reviewed and approved by our university Institutional Review Board (IRB). The study consisted of one online Qualtrics survey that lasted an average of 14 min. Each participant first completed a background survey that captured their self-identified demographics, including race, gender, and education. Afterwards, they were asked to familiarize themselves with the racial terms as defined by the U.S. Census Bureau research (Matthews et al., 2017). Participants were then asked to categorize the 42 avatars by their perceived race and gender. Participants were shown only one avatar at a time and the order was randomized.

For each of the avatars, participants were shown three perspectives: a 45° left headshot, a direct or 0° headshot, and a 45° right headshot (see Figure 3). Avatars were shown from the shoulders up and were dressed in a plain gray shirt. The images were rendered in Unity using the standard diffuse shader and renderer. The avatars were illuminated by a soft white (#FFFEF5) directional light with an intensity of 1.0, and light gray (#7F7F7F) was used for the background. Participants were asked to select all races that each avatar could represent: “American Indian or Alaskan Native”, “Asian”, “Black or African American”, “Hispanic, Latino, or Spanish”, “Middle Eastern or North African”, “Native Hawaiian or Pacific Islander”, “White”, or “Other”. “Other” included an optional textbox if a participant wanted to be specific. We allowed participants to select multiple categories according to recommendations for surveying race (Matthews et al., 2017). For gender, participants were able to select “Male”, “Female”, or “Non-binary”. Participants were paid \$5.00 via Prolific for completing the study.

### 4.2 Participants

A total of 132 participants (65 male, 63 female, 4 non-binary) from 33 different countries were recruited to take part in the study. We aimed to ensure a diverse representation of perspectives by balancing participants by race, gender, and country. Despite multiple recruitment attempts, including targeted solicitations via Prolific, we had difficulty recruiting NHPI participants. It is important to note that we excluded volunteers who had previously assisted with modeling the avatars from participating in the validation study to avoid potentially overfitting their own biases.

### 4.3 Data analysis and labeling approach

To validate the racial identification of our virtual avatars, we used Cochran’s Q test (Sheskin, 2011), which allowed us to

**Validation**



**FIGURE 3**  
An example of how each avatar was presented to participants during our validation study.

**TABLE 2 Breakdown of our validation study's participants by ethnicity, mean age (and standard deviation), gender, and country.**

Ethnicity	Age	Gender	Country
AIAN	37.4 (15.5)	9M, 10F, 1NB	U.S. (14), Canada (3), Chile (1), Mexico (1), Cambodia (1)
Asian	28.5 (9.4)	10M, 10F	United Kingdom (5), Canada (3), South Africa (3), India (2),
			Indonesia (2), China (1), France (1), Germany (1),
			Italy (1), Malaysia (1)
Black	23.4 (3.6)	10M, 9F, 1NB	South Africa (16), Nigeria (2), Swaziland (1), United Kingdom (1)
Hispanic	23.8 (3.2)	10M, 9F, 1NB	Mexico (15), Chile (3), Portugal (2)
MENA	30.6 (10.2)	10M, 10F	Israel (9), Lebanon (3), Jordan (2), Bahrain (1), Egypt (1),
			Iran (1), Iraq (1), Saudi Arabia (1),
			Syrian Arabic Republic (1)
NHPI	33.7 (9.9)	7M, 5F	New Zealand (Maori) (8), Samoa (2), Fiji (1), U.S. (1)
White	25.4 (6.1)	9M, 10F, 1NB	Portugal (5), Italy (5), Poland (3), Mexico (2), Belgium (1),
			Greece (1), Ireland (1), United Kingdom (1), U.S. (1)

analyze any significant differences among the selected race categories. This approach was necessary since our survey format allowed participants to select more than one race category for each avatar, following the U.S. Census Bureau's research recommendations (Matthews et al., 2017). Since the Chi-squared goodness of fit test requires mutually exclusive categories, we were unable to use it in our analysis. Furthermore, since our data was dichotomous, a repeated-measures analysis of variance (ANOVA) was not appropriate. Therefore, Cochran's Q test was the most appropriate statistical analysis method for our survey data.

We used a rigorous statistical approach to assign race and gender labels to each avatar. First, we conducted the Cochran's Q test across all participants ( $n = 132$ ) at a 95% confidence level to identify significant differences in the participants' responses. If the test indicated significant differences, we performed pairwise comparisons between each race using Dunn's test to determine which races were significantly different.

For each avatar, we assigned a race label if the race was selected by the majority of participants (i.e., over 50% of participants selected it) and if the race was selected significantly more than other race choices and not significantly less than any other race. This approach resulted in a single race label for most avatars, but some avatars were assigned multiple race labels due to multiple races being selected significantly more than all other races. If no race was selected significantly more than the majority, then we categorized the avatar as "Ambiguous". We followed a similar procedure for assigning gender labels.

To account for the possibility that the race of the participant might influence their perception of virtual race, we also assigned labels based on same-race participants. This involved using the same procedure for assigning labels as described above, except based only on the selections of participants who identified as the same race as the avatar. This also allows future researchers to have the flexibility to use the labels from all study participants for studies focused on individuals from diverse racial backgrounds or to use the labels from participants of the same race for studies targeting specific racial groups.

**TABLE 3** Assigned labels for all 42 base avatars. “All” indicates that the label was identified by all 132 participants, while “Same-Race” only includes the data of participants who identify as the race that the avatar was modeled for. Agreement labels were calculated as the percentage of participants who perceived an avatar to represent a race or gender.

Avatar	Race (all)	Agreement	Race (same-race)	Agreement	Gender (all)	Agreement
AIAN_M_1	AIAN	0.67	AIAN	0.90	Male	0.89
AIAN_M_2	AIAN	0.65	AIAN	0.75	Male	0.89
AIAN_F_1	AIAN	0.55	AIAN	0.60	Female	0.64
AIAN_F_2	AIAN	0.55	AIAN	0.75	Female	0.75
Asian_M_1	Asian	0.86	Asian	0.90	Male	0.95
Asian_M_2	Asian	0.98	Asian	1.00	Male	0.91
Asian_M_3	Asian	0.98	Asian	1.00	Male	0.91
Asian_F_1	Asian	0.98	Asian	1.00	Female	0.98
Asian_F_2	Asian	0.93	Asian	0.90	Female	0.96
Asian_F_3	Asian	0.95	Asian	1.00	Female	0.97
Black_M_1	Black	0.98	Black	1.00	Male	0.98
Black_M_2	Black	0.97	Black	1.00	Male	0.98
Black_M_3	Black	0.98	Black	1.00	Male	0.73
Black_F_1	Black	0.98	Black	1.00	Female	0.98
Black_F_2	Black	0.98	Black	1.00	Female	0.94
Black_F_3	Black	0.95	Black	1.00	Female	0.97
Hispanic_M_1	Hispanic/White	0.65/0.55	Hispanic	0.85	Male	0.98
Hispanic_M_2	White	0.64	Hispanic	0.80	Male	0.98
Hispanic_M_3	Hispanic/White	0.55/0.51	Hispanic	0.75	Male	0.97
Hispanic_F_1	Hispanic	0.64	Hispanic	0.80	Female	0.95
Hispanic_F_2	White	0.70	Hispanic/White	0.55/0.60	Female	0.95
Hispanic_F_3	Hispanic/White	0.59/0.52	Hispanic	0.75	Female	0.93
MENA_M_1	White	0.56	MENA	0.70	Male	0.99
MENA_M_2	White	0.64	MENA/White	0.65/0.60	Male	1.00
MENA_M_3	MENA/White	0.55/0.65	MENA/White	0.75/0.55	Male	0.98
MENA_F_1	White	0.58	MENA/White	0.70/0.60	Female	0.98
MENA_F_2	White	0.60	MENA	0.60	Female	0.98
NHPI_M_1	NHPI	0.52	NHPI	0.58	Male	0.98
NHPI_M_2	Hispanic	0.65	NHPI	0.67	Male	1.00
White_M_1	White	0.96	White	0.95	Male	0.99
White_M_2	White	0.98	White	0.95	Male	1.00
White_M_3	White	0.93	White	0.90	Male	0.99
White_F_1	White	0.94	White	0.95	Female	0.97
White_F_2	White	0.96	White	0.95	Female	0.98
White_F_3	White	0.94	White	0.95	Female	0.99
X_AIAN_M_1	Hispanic	0.64	Hispanic	0.75	Male	0.99
X_AIAN_F_1	Hispanic	0.54	Hispanic	0.45	Female	0.86

(Continued on following page)



**TABLE 3 (Continued)** Assigned labels for all 42 base avatars. “All” indicates that the label was identified by all 132 participants, while “Same-Race” only includes the data of participants who identify as the race that the avatar was modeled for. Agreement labels were calculated as the percentage of participants who perceived an avatar to represent a race or gender.

Avatar	Race (all)	Agreement	Race (same-race)	Agreement	Gender (all)	Agreement
X_MENA_F_1	Ambiguous	N/A	Ambiguous	N/A	Female	0.98
X_NHPI_M_1	Hispanic	0.55	Asian	0.58	Male	0.98
X_NHPI_F_1	Hispanic	0.55	Ambiguous	N/A	Female	0.92
X_NHPI_F_2	Hispanic	0.58	Ambiguous	N/A	Female	0.95
X_NHPI_F_3	NHPI	0.52	Ambiguous	N/A	Female	0.92

## 5 Results

### 5.1 Validated avatar labels

Table 3 summarizes our results and labels for all 42 base avatars across all participants and for same-race participants.

#### 5.1.1 Race and gender labels

Asian, Black, and White avatars were correctly identified as their intended race across all participants, while most of the remaining avatars were accurately identified by same-race participants (see Table 3 for all and same-race agreement rates). Therefore, we observed some differences in identification rates based on the race of the participants, highlighting the potential impact of own-race bias on the perception of virtual avatars. Notably, there were no significant differences in gender identification rates based on participant race, indicating that all avatars were correctly perceived as their intended gender by all participants, regardless of their racial background.

#### 5.1.2 Naming convention

If an avatar was identified as its intended race by corresponding same-race participants, we named it after that race. For instance, the avatar *Hispanic\_M\_2* was labeled as White by all participants. However, our Hispanic participants perceived it as solely Hispanic. Hence, we left the original name. However, if an avatar was labeled as “Ambiguous” or as a different race by same-race participants, we added an X at the beginning of its name to indicate that it was not validated. Avatars were also labeled by their identified gender (“M” or “F”).

### 5.2 Other-race vs. same-race perception

To further examine how participant race affected perception of virtual avatar race, we additionally analyzed the data by separating same-race and other-race agreement rates. In effect, we separated the selections of the participants who were the same race as the avatar modeled and those who were not.

#### 5.2.1 Difference in agreement rates

Figure 4 displays the difference in agreement rates between same-race and other-race. Figure 4 shows that several avatars were strongly identified by both other-race and same-race participants. In particular, all Asian, Black, and White avatars were perceived as

their intended race with high agreement rates by both same-race and other-race participants (over 90% agreement for all but one). However, some avatars were only identified by participants of the same race as the avatar. For example, our analysis of the agreement rates for different racial groups revealed interesting trends. For instance, non-Hispanic participants had an average agreement rate of 54.5% for Hispanic avatars, while Hispanic participants had a much higher average agreement of 75.0%. Similar patterns were observed for AIAN (57.8% other-race, 75.0% same-race) and MENA (40.4% other-race, 68.0% same-race) avatars.

#### 5.2.2 Perceived race clusters

To gain deeper insights into how participants perceived the avatars’ races, we employed Principle Component Analysis (PCA) to reduce the agreement rates of each of the 42 base avatars down to two dimensions. Next, we performed K-means clustering (Likas et al., 2003) on the resulting two-dimensional data to group the avatars based on their perceived race. We optimized the number of clusters using the elbow method and distortion scores (Bengfort et al., 2018). We applied this technique to both other-race and same-race agreement rates to determine whether there were any differences in the clustering based on participant race. By visualizing the clusters, we aimed to better understand the differences in how participants perceived the avatars’ races.

Figure 5 shows that Asian, Black, and White avatars were perceived consistently by all participants, with clearly defined clusters. However, there was more confusion in perceiving AIAN, Hispanic, MENA, and NHPI avatars, which clustered closer together. Same-race participants had less overlap and more-accurately perceived these avatars, with more separation between them. For example, the Hispanic and MENA avatars were in separate clusters for same-race participants, except for one avatar (*Hispanic\_F\_2*). On the other hand, the Hispanic and MENA avatars were entirely clustered together for other-race participants.

## 6 Discussion

In this section, we discuss the validation of our avatars. Specifically, we examine the extent to which each avatar was correctly identified as its intended race and the variability in identification across different participant groups. Additionally, we discuss the implications of our results for virtual avatar research, highlighting the importance of considering the potential impact of own-race bias on avatar race perception. Finally, we describe the

Avatar	Other-Race Participants							Same-Race Participants						
	AIAN	Asian	Black	Hispanic	MENA	NHPI	White	AIAN	Asian	Black	Hispanic	MENA	NHPI	White
AIAN_M_1	0.63	0.04	0.09	0.40	0.11	0.46	0.12	0.90	0.05	0.05	0.20	0.15	0.10	0.05
AIAN_M_2	0.63	0.03	0.07	0.45	0.12	0.41	0.09	0.75	0.00	0.00	0.30	0.10	0.40	0.10
AIAN_F_1	0.54	0.08	0.04	0.48	0.20	0.35	0.21	0.60	0.00	0.00	0.40	0.10	0.25	0.15
AIAN_F_2	0.51	0.09	0.08	0.53	0.23	0.31	0.21	0.75	0.05	0.00	0.15	0.30	0.25	0.15
Asian_M_1	0.12	0.85	0.00	0.12	0.03	0.19	0.21	0.15	0.90	0.00	0.05	0.00	0.15	0.10
Asian_M_2	0.11	0.97	0.00	0.05	0.02	0.10	0.07	0.05	1.00	0.00	0.00	0.00	0.00	0.10
Asian_M_3	0.05	0.98	0.00	0.04	0.01	0.12	0.11	0.00	1.00	0.00	0.00	0.00	0.10	0.15
Asian_F_1	0.08	0.97	0.00	0.04	0.03	0.12	0.06	0.00	1.00	0.05	0.00	0.00	0.05	0.05
Asian_F_2	0.12	0.94	0.01	0.04	0.04	0.18	0.05	0.20	0.90	0.00	0.00	0.05	0.05	0.05
Asian_F_3	0.18	0.95	0.00	0.06	0.03	0.13	0.07	0.05	1.00	0.05	0.05	0.05	0.05	0.00
Black_M_1	0.02	0.04	0.97	0.05	0.15	0.12	0.02	0.00	0.00	1.00	0.00	0.25	0.05	0.00
Black_M_2	0.03	0.01	0.96	0.05	0.20	0.04	0.00	0.00	0.00	1.00	0.05	0.15	0.00	0.00
Black_M_3	0.03	0.02	0.98	0.04	0.18	0.04	0.00	0.00	0.00	0.95	0.00	0.25	0.00	0.00
Black_F_1	0.01	0.02	0.98	0.04	0.17	0.06	0.00	0.00	0.00	1.00	0.05	0.25	0.00	0.00
Black_F_2	0.03	0.02	0.97	0.05	0.12	0.07	0.00	0.00	0.00	1.00	0.05	0.25	0.00	0.00
Black_F_3	0.03	0.02	0.94	0.07	0.16	0.08	0.00	0.00	0.00	1.00	0.05	0.10	0.05	0.00
Hispanic_M_1	0.11	0.05	0.03	0.60	0.29	0.12	0.58	0.25	0.15	0.00	0.85	0.25	0.10	0.40
Hispanic_M_2	0.12	0.06	0.02	0.46	0.19	0.12	0.71	0.25	0.00	0.00	0.80	0.25	0.20	0.30
Hispanic_M_3	0.14	0.08	0.03	0.52	0.26	0.14	0.54	0.10	0.10	0.00	0.75	0.10	0.20	0.30
Hispanic_F_1	0.13	0.09	0.04	0.61	0.31	0.14	0.43	0.15	0.00	0.05	0.80	0.30	0.10	0.15
Hispanic_F_2	0.08	0.05	0.04	0.52	0.15	0.09	0.71	0.10	0.10	0.00	0.55	0.30	0.05	0.60
Hispanic_F_3	0.23	0.08	0.05	0.56	0.22	0.09	0.55	0.15	0.05	0.00	0.75	0.25	0.15	0.30
MENA_M_1	0.09	0.03	0.00	0.50	0.46	0.05	0.59	0.10	0.05	0.00	0.25	0.70	0.05	0.40
MENA_M_2	0.06	0.04	0.00	0.34	0.47	0.07	0.64	0.10	0.05	0.00	0.25	0.65	0.10	0.60
MENA_M_3	0.06	0.02	0.01	0.38	0.51	0.05	0.67	0.10	0.05	0.00	0.45	0.75	0.05	0.55
MENA_F_1	0.12	0.09	0.02	0.49	0.31	0.11	0.58	0.10	0.15	0.00	0.25	0.70	0.10	0.60
MENA_F_2	0.20	0.07	0.02	0.41	0.27	0.10	0.62	0.15	0.10	0.00	0.40	0.60	0.10	0.45
NHPI_M_1	0.28	0.34	0.07	0.45	0.16	0.52	0.08	0.25	0.42	0.00	0.33	0.00	0.58	0.00
NHPI_M_2	0.28	0.17	0.08	0.66	0.27	0.46	0.15	0.17	0.25	0.00	0.42	0.08	0.67	0.00
White_M_1	0.03	0.00	0.00	0.08	0.10	0.03	0.96	0.05	0.00	0.00	0.05	0.00	0.05	0.95
White_M_2	0.05	0.01	0.00	0.10	0.08	0.04	0.98	0.05	0.00	0.00	0.05	0.10	0.10	0.95
White_M_3	0.04	0.02	0.00	0.15	0.09	0.04	0.94	0.00	0.00	0.00	0.05	0.10	0.10	0.90
White_F_1	0.06	0.04	0.00	0.17	0.06	0.04	0.94	0.10	0.00	0.05	0.05	0.00	0.10	0.95
White_F_2	0.05	0.04	0.00	0.09	0.06	0.03	0.96	0.05	0.05	0.00	0.00	0.00	0.00	0.95
White_F_3	0.08	0.01	0.00	0.19	0.05	0.05	0.94	0.00	0.00	0.00	0.15	0.00	0.05	0.95
X_AIAN_M_1	0.22	0.16	0.04	0.62	0.29	0.25	0.33	0.15	0.10	0.10	0.75	0.15	0.25	0.35
X_AIAN_F_1	0.30	0.24	0.03	0.55	0.21	0.27	0.27	0.25	0.15	0.00	0.45	0.15	0.25	0.35
X_MENA_F_1	0.14	0.08	0.04	0.43	0.33	0.08	0.48	0.20	0.10	0.00	0.40	0.45	0.05	0.50
X_NHPI_M_1	0.33	0.25	0.03	0.60	0.23	0.56	0.11	0.25	0.58	0.00	0.33	0.00	0.42	0.00
X_NHPI_F_1	0.29	0.24	0.34	0.56	0.20	0.42	0.05	0.17	0.17	0.25	0.50	0.08	0.42	0.25
X_NHPI_F_2	0.30	0.19	0.30	0.59	0.28	0.38	0.08	0.25	0.00	0.33	0.42	0.25	0.50	0.08
X_NHPI_F_3	0.33	0.33	0.14	0.49	0.17	0.54	0.05	0.42	0.08	0.00	0.42	0.00	0.42	0.00

**FIGURE 4** Confusion matrix heatmap of agreement rates for the 42 base avatars by separated by other-race participants and same-race participants (i.e., participants of a different or same race as the avatar). Agreement rates were calculated as the percentage of participants who perceived an avatar to represent a race or gender.

potential future impact of our avatar library in the community, including how it can be used to promote diversity and inclusion.

## 6.1 Race identification

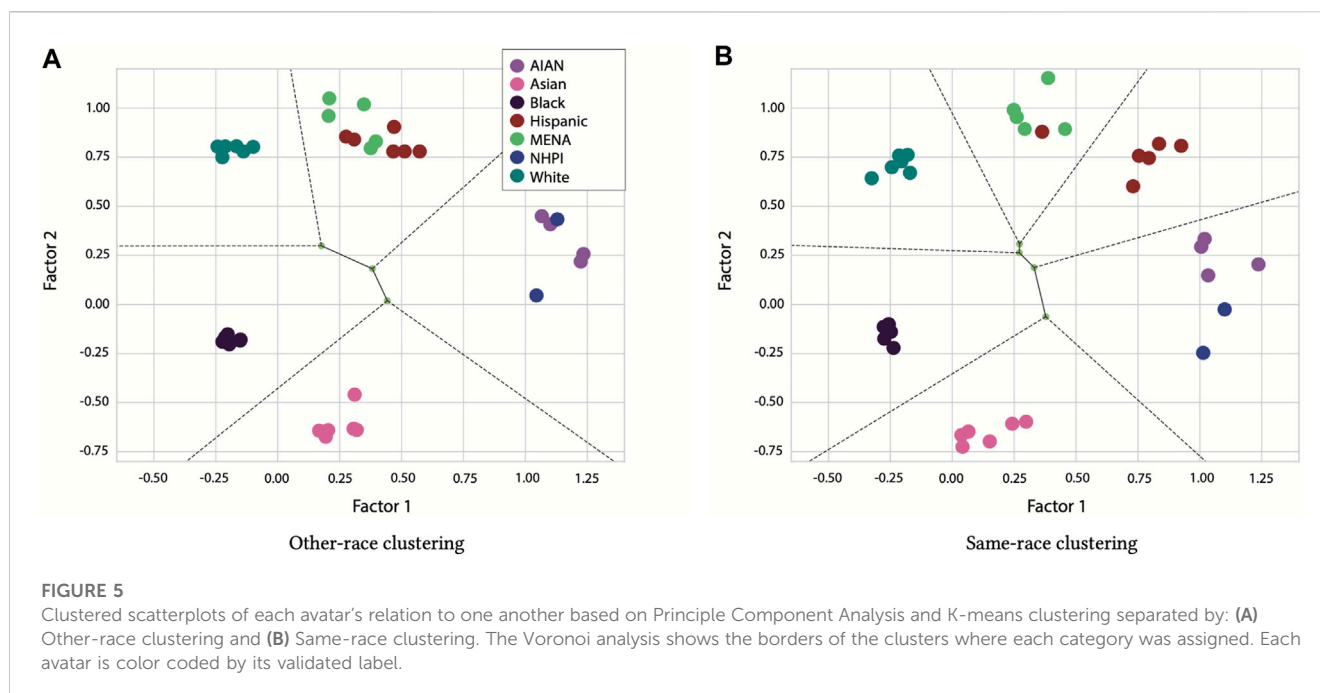
### 6.1.1 Universally identified avatars

We found that our Asian, Black, and White avatars were recognized by all participants with high agreement rates. The high agreement rates observed across all participants are particularly remarkable given prior research demonstrating that people are less accurate when identifying faces of races other than their own, particularly in ethnically homogeneous countries (Anzures et al., 2013). Despite this psychological phenomenon, our participants, who equitably represented different races, consistently provided high agreement rates for the avatars. This suggests that these avatars can be a valuable tool for researchers seeking to create virtual humans that can be easily identified by individuals from different racial backgrounds.

Our results may be due to perceptual expertise or familiarity with other-race faces, as proposed by Civile and McLaren (2022). We hypothesize that this familiarity could be explained by the prevalence of these racial groups in global media and pop culture. For example, White cast members were the most represented in popular Hollywood movies over the last decade, followed by Black cast members (Malik et al., 2022). Since Hollywood movies have a dominant share in the global film industry (Maisuwong, 2012), people may be more familiar with characters that are prevalent in these films. Additionally, East Asian media culture has become widely popular worldwide over the past few decades (Iwabuchi, 2010; Jin et al., 2021). Phenomena like “The Hallyu Wave” and “Cool Japan” (Lux, 2021) have enabled East Asian films, dramas, and pop music to gain a global following. As people may often encounter these racial groups in media, this familiarity may have facilitated their recognition of these avatars.

### 6.1.2 Same-race identified avatars

As expected, some avatars were only identified by participants of the same race as the avatar, consistent with the own-race bias effect.



For example, as seen in [Table 3](#), the Hispanic avatars received mixed ratings of White and Hispanic across all participants, but most were perceived as solely Hispanic by Hispanic-only participants. Similarly, only one MENA avatar was perceived as MENA by all participants, while five were perceived as MENA by MENA-only participants. These results suggest that participants' own-race bias, a well-known phenomenon in psychology, may also affect their perception of virtual avatars. The findings point to the importance of considering participants' race when using virtual avatars in research or applications that require accurate representation of different racial groups.

### 6.1.3 Ambiguous avatars

Several avatars in our library were perceived ambiguously by all participants and only same-race participants, and therefore labeled as such (see [Table 3](#) for details). Identifying the reason for these avatars' lack of clear identification is not straightforward, and multiple factors could be at play. For instance, the two ambiguous AIAN avatars were the only ones with short hairstyles, which may have impacted their identification as AIAN. Long hair carries cultural and spiritual significance in many AIAN tribes ([Thanikachalam et al., 2019](#)), and some participants may have perceived the avatars as non-AIAN as a result, even among AIAN participants.

The validation of our NHPI avatars was limited, possibly due to the low number of NHPI participants ( $n = 12$ ) in our study, despite our targeted recruitment efforts. As a consequence, most of the NHPI avatars were not validated by NHPI participants, including the lack of validation for any female NHPI avatars. Another potential reason for this lack of validation is that the majority of our NHPI participants identified themselves as New Zealand Maori, whereas our avatars were developed with the help of Samoan and Native Hawaiian volunteer representatives. Therefore, it is possible that our NHPI avatars are representative of some NHPI cultures, but

not New Zealand Maori. In future studies, expanding recruitment efforts for both interview volunteers and study participants will be crucial, despite the challenges involved in doing so. For example, future studies may need to compensate NHPI participants more than participants of other races.

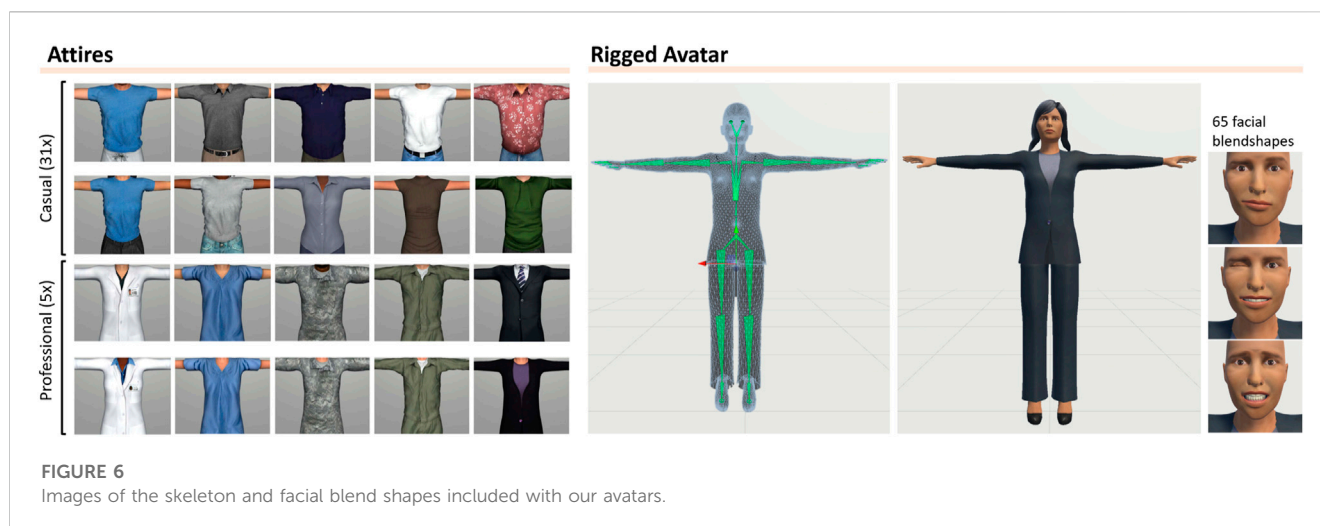
## 6.2 Implications for virtual avatars

Our study provides valuable insights for virtual avatar applications and research. Our findings indicate that human behavior in race categorization can apply to virtual avatars, which has notable implications for interactions in virtual experiences. [Kawakami et al. \(2017\)](#) suggest that in-group and out-group categorization can lead to stereotyping, social judgments, and group-based evaluations. Therefore, designers and developers should be aware of this and take necessary steps to mitigate unintended consequences in virtual experiences. For example, regulating codes of conduct ([Taylor et al., 2020](#)) can help to improve interracial interactions in VR.

Interestingly, our study also replicated a nuanced finding from more recent psychology research on the perception of ambiguous avatars ([Nicolas et al., 2019](#)). As seen in [Table 3](#), most of the misidentified avatars were identified as Hispanic by all participants. Similarly, [Nicolas et al. \(2019\)](#) recently found that participants classify racially ambiguous photos as Hispanic or MENA, regardless of their parent ethnicities. We believe that this effect extended to our virtual avatars.

## 6.3 An open library of validated avatars

As a contribution to the research community, we are providing open access to our virtual avatar library, which includes all 210 fully rigged avatars, along with validated labels for each avatar's race and



**FIGURE 6**  
Images of the skeleton and facial blend shapes included with our avatars.

gender. Our library features avatars of seven different races, providing a diverse selection for researchers to use in their studies. The validated labels can facilitate research on the impact of avatar race, and researchers can choose to use the labels for studies aimed at individuals from different racial backgrounds or same-race labels for specific study populations.

The *Virtual Avatar Library for Inclusion and Diversity (VALID)* provides researchers and developers with a diverse set of fully rigged avatars suitable for various scenarios such as casual, business, medical, military, and utility. Each avatar comes with 65 facial blend shapes, enabling dynamic facial expressions (see Figure 6). The library is readily available for download and can be used in popular game engines like Unity or Unreal. Although this is the first iteration of the library, we plan to update it by adding more professions, genders, and outfits soon. In addition, the library can be used for a wide range of research purposes, including social psychology simulations and educational applications.

## 6.4 Limitations

We recognize that our VALID avatar library is only a small step towards achieving greater diversity and inclusion in avatar resources. We acknowledge that the representation of each demographic is limited and plan to expand the diversity within each group by creating new avatars. For example, our Asian avatars are modeled entirely after East Asians, but we plan to expand VALID to introduce different regional categories, such as South Asian and Southeast Asian avatars. Our Hispanic representatives have pointed out the need for more diverse Hispanic avatars, including varying skin tones to represent different South American populations, such as Mexican and Cuban. Our NHPI representatives have suggested the inclusion of tattoos, which hold cultural significance for some NHPI communities, could improve the identifiability of our NHPI avatars, in addition to improving our NHPI recruitment methods. Any future updates to the library will undergo the same rigorous creation, iteration, and validation process as the current avatars.

While our first iteration of the library focused on diversity in terms of race, we realize that the avatars mostly represent young and fit adults, which does not reflect all types of people. In the future, we

plan to update the library with a diversity of body types that include different body mass index (BMI) representations and ages. Including avatars with different BMI representations is not only more inclusive, but can also be useful for studies targeting physical activity, food regulation, and therapy (Scarpina et al., 2019). Likewise, we plan to include shaders and bump maps (Hughes, 2011) that can age any given avatar by creating realistic wrinkles and skin folds, further improving the diversity and inclusivity of VALID.

Another limitation of the current work is that our library includes only male and female representations. In future updates, we plan to include non-binary and androgynous avatars. Currently, there are not many androgynous models that are freely available. However, they can be an area of important study. For example, previous studies found that androgynous avatars reduce gender bias and stereotypical assumptions in virtual agents (Nag and Yalçın, 2020) and improve student attitudes (Booth et al., 2010). Thus, we plan to include these avatars in a future update by following Nag and Yalçın (2020)'s guidelines for creating androgynous virtual humans.

Our study, while diverse in terms of race and country, is not representative of everyone. We recruited participants through the online platform Prolific, which is known for its increased diversity compared to other crowdsourcing platforms such as Mechanical Turk. However, due to the online nature of the platform, we primarily recruited younger adults. It is possible that perceptions of our avatars may differ among other age groups, such as children or older adults. Therefore, it is important to broaden recruitment efforts by exploring alternative platforms and recruitment strategies that may be more effective in reaching a wider range of participants. Future studies could also consider conducting in-person studies or focus groups to gather additional insights into avatar perception.

## 7 Conclusion

We have introduced a new virtual avatar library comprised of 210 fully rigged avatars with diverse professions and outfits, available for free. Our library aims to promote diversity and inclusion by creating equitable representation of seven races across various professions. We designed 42 base avatars using data-driven facial averages and collaborated with volunteer



representatives of each ethnicity. A large validation study involving participants from around the world was conducted to obtain validated labels and metadata for the perceived race and gender of each avatar. Additionally, we offer a comprehensive process for creating, iterating, and validating diverse avatars to aid other researchers in creating similarly validated avatars.

Our validation study revealed that the majority of avatars were accurately perceived as the race they were modeled for. However, we observed that some avatars, such as the Hispanic and MENA avatars, were only validated as such by participants who identified as Hispanic or MENA, respectively. This finding suggests that the perception of virtual avatars may be influenced by own-race bias or the other-race effect, as described in the psychology literature. Moving forward, we plan to expand the library to include additional races, professions, body types, age ranges, and gender representations to further improve diversity and inclusion.

## Data availability statement

The original contributions presented in the study are publicly available. This data can be sourced at <https://github.com/xrtilab/Validated-Avatar-Library-for-Inclusion-and-Diversity—VALID>.

## Ethics statement

The studies involving humans were approved by the University of Central Florida Institutional Review Board. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## References

- Ali, M. R., Razavi, S. Z., Langevin, R., Al Mamun, A., Kane, B., Rawassizadeh, R., et al. (2020). "A virtual conversational agent for teens with autism spectrum disorder: experimental results and design lessons," in Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents, Virtual Event Scotland, UK, October, 2020. doi:10.1145/3383652.3423900
- Anzures, G., Quinn, P. C., Pascalis, O., Slater, A. M., Tanaka, J. W., and Lee, K. (2013). Developmental origins of the other-race effect. *Curr. Dir. Psychol. Sci.* 22, 173–178. doi:10.1177/0963721412474459
- Bainbridge, W. A., Isola, P., and Oliva, A. (2013). The intrinsic memorability of face photographs. *J. Exp. Psychol. General* 142, 1323–1334. doi:10.1037/a0033872
- Baylor, A. (2003). Effects of images and animation on agent persona. *J. Educ. Comput. Res.* 28, 373–394.
- Baylor, A. L. (2009). Promoting motivation with virtual agents and avatars: role of visual presence and appearance. *Philosophical Trans. R. Soc. B Biol. Sci.* 364, 3559–3565. doi:10.1098/rstb.2009.0148
- Bengfort, B., Bilbro, R., Danielsen, N., Gray, L., McIntyre, K., Roman, P., et al. (2018). *Yellowbrick*. doi:10.5281/zenodo.1206264
- Bickmore, T., Parmar, D., Kimani, E., and Olafsson, S. (2021). "Diversity informatics: reducing racial and gender bias with virtual agents," in Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents, Virtual Event Japan, September, 2021, 25–32. doi:10.1145/3472306.3478365
- Blascovich, J., Wyer, N. A., Swart, L. A., and Kibler, J. L. (1997). Racism and racial categorization. *J. Personality Soc. Psychol.* 72, 1364–1372. doi:10.1037/0022-3514.72.6.1364
- Blom, K. J., Bellido Rivas, A. I., Alvarez, X., Cetinaslan, O., Oliveira, B., Orvalho, V., et al. (2014). Achieving participant acceptance of their avatars. *Presence Teleoperators Virtual Environ.* 23, 287–299. doi:10.1162/PRES\_a\_00194
- Boberg, M., Piippo, P., and Ollila, E. (2008). "Designing avatars," in Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts (ACM), Athens, Greece, September, 2008, 232–239. doi:10.1145/1413634.1413679
- Booth, S., Goodman, S., Kirkup, G., Gulz, A., and Haake, M. (2010). *Challenging gender stereotypes using virtual pedagogical characters*, Pennsylvania, PA, United States: IGI Global, 113–132. doi:10.4018/978-1-61520-813-5.ch007
- Butler, C., Michalowicz, S., Subramanian, L., and Burleson, W. (2017). "More than a feeling: the miface framework for defining facial communication mappings," in Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, Québec City QC, Canada, October, 2017, 773–786. doi:10.1145/3126594.3126640
- Chapman, E. N., Kaatz, A., and Carnes, M. (2013). Physicians and implicit bias: how doctors may unwittingly perpetuate health care disparities. *J. general Intern. Med.* 28, 1504–1510. doi:10.1007/s11606-013-2441-1
- Cheng, Y. F., Yin, H., Yan, Y., Gugenheimer, J., and Lindlbauer, D. (2022). "Towards understanding diminished reality," in Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, New Orleans LA, USA, April, 2022. doi:10.1145/3491102.3517452
- Civile, C., and McLaren, I. P. L. (2022). Transcranial direct current stimulation (tDCS) eliminates the other-race effect (ORE) indexed by the face inversion effect for own versus other-race faces. *Sci. Rep.* 12, 1–10. doi:10.1038/s41598-022-17294-w
- DeBruine, L. (2018). *debruine/webmorph: beta release 2*. <https://zenodo.org/records/1162670>.
- Dewez, D., Hoyet, L., Lécuyer, A., and Argelaguet Sanz, F. (2021). "Towards 'avatar-friendly' 3d manipulation techniques: bridging the gap between sense of embodiment and interaction in virtual reality," in Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, May, 2021. doi:10.1145/3411764.3445379
- Fox, J., Ahn, S. J. G., Janssen, J. H., Yeykelis, L., Segovia, K. Y., and Bailenson, J. N. (2015). Avatars versus agents: a meta-analysis quantifying the effect of agency on social influence. *Human-Computer Interact.* 30, 401–432. doi:10.1080/07370024.2014.921494

## Author contributions

TD, SZ, RM, and MG-F contributed to conception and design of the library and validation user study. TD and SZ conducted design interviews. TD created the avatars and organized the library repository. TD performed the statistical analysis and visualizations. TD wrote the first draft of the manuscript. TD, RM, and MG-F wrote sections of the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

Internal funding from the University of Central Florida was used to support this research.

## Conflict of interest

MG-F was employed by the company Google.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.



- Gamberini, L., Chittaro, L., Spagnoli, A., and Carlesso, C. (2015). Psychological response to an emergency in virtual reality: effects of victim ethnicity and emergency type on helping behavior and navigation. *Comput. Hum. Behav.* 48, 104–113. doi:10.1016/j.chb.2015.01.040
- Ge, L., Zhang, H., Wang, Z., Quinn, P. C., Pascalis, O., Kelly, D., et al. (2009). Two faces of the other-race effect: recognition and categorisation of Caucasian and Chinese faces. *Perception* 38, 1199–1210. doi:10.1068/p6136
- Gilbert, J. E., Arbutnot, K., Hood, S., Grant, M. M., West, M. L., Mcmillan, Y., et al. (2008). Teaching algebra using culturally relevant virtual instructors. *Int. J. Virtual Real.* 7, 21–30.
- Gonzalez-Franco, M., Bellido, A. I., Blom, K. J., Slater, M., and Rodriguez-Fornells, A. (2016). The neurological traces of look-alike avatars. *Front. Hum. Neurosci.* 10, 392. doi:10.3389/fnhum.2016.00392
- Gonzalez-Franco, M., Egan, Z., Peachey, M., Antley, A., Randhavane, T., Panda, P., et al. (2020a). Movebox: democratizing mocap for the microsoft rocketbox avatar library. In 2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR). Utrecht, Netherlands, December, 2020, 91–98. doi:10.1109/AIVR50618.2020.00026
- Gonzalez-Franco, M., Ofek, E., Pan, Y., Antley, A., Steed, A., Spanlang, B., et al. (2020b). The rocketbox library and the utility of freely available rigged avatars. *Front. Virtual Real.* 1. doi:10.3389/frvir.2020.561558
- Gonzalez-Franco, M., Slater, M., Birney, M. E., Swapp, D., Haslam, S. A., and Reicher, S. D. (2018). Participant concerns for the learner in a virtual reality replication of the milgram obedience study. *PLoS one* 13, e0209704. doi:10.1371/journal.pone.0209704
- Groom, V., Bailenson, J. N., and Nass, C. (2009). The influence of racial embodiment on racial bias in immersive virtual environments. *Soc. Infl.* 4, 231–248. doi:10.1080/15534510802643750
- Halan, S., Sia, I., Cray, M., and Lok, B. (2015). Exploring the effects of healthcare students creating virtual patients for empathy training. In Intelligent Virtual Agents: 15th International Conference, IVA 2015, Delft, Netherlands, August, 2015, 239–249.
- Heidicker, P., Langbehn, E., and Steinicke, F. (2017). “Influence of avatar appearance on presence in social vr,” in 2017 IEEE Symposium on 3D User Interfaces (3DUI), Los Angeles, CA, USA, March, 2017, 233–234. doi:10.1109/3DUI.2017.7893357
- Huang, A., Knierim, P., Chioffi, F., Chuang, L. L., and Welsch, R. (2022). “Proxemics for human-agent interaction in augmented reality,” in Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, New Orleans LA USA, April, 2022. doi:10.1145/3491102.3517593
- Hughes, R. (2011). Real-time wrinkles for expressive virtual characters. Ph.D. thesis. Dublin, Ireland: Trinity College Dublin.
- Iwabuchi, K. (2010). Globalization, East Asian media cultures and their publics. *Asian J. Commun.* 20, 197–212. doi:10.1080/01292981003693385
- Jeong, D. C., Xu, J. J., and Miller, L. C. (2020). “Inverse kinematics and temporal convolutional networks for sequential pose analysis in vr,” in 2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), Utrecht, Netherlands, December, 2020, 274–281. doi:10.1109/AIVR50618.2020.00056
- Jin, D., Yoon, K., and Min, W. (2021). “Transnational Hallyu: the globalization of Korean digital and popular culture,” in *Asian cultural studies: transnational and dialogic approaches* (Lanham, Maryland, USA: Rowman and Littlefield Publishers).
- Jo, D., Kim, K., Welch, G. F., Jeon, W., Kim, Y., Kim, K.-H., et al. (2017). “The impact of avatar-owner visual similarity on body ownership in immersive virtual reality,” in Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology, Gothenburg, Sweden, November, 2017. doi:10.1145/3139131.3141214
- Jung, S., Bruder, G., Wisniewski, P. J., Sandor, C., and Hughes, C. E. (2018). “Over my hand: using a personalized hand in vr to improve object size estimation, body ownership, and presence,” in Proceedings of the Symposium on Spatial User Interaction, Berlin, Germany, October, 2018, 60–68. doi:10.1145/3267782.3267920
- Kawakami, K., Amodio, D., and Hugenberg, K. (2017). “Chapter one - intergroup perception and cognition,” in *An integrative framework for understanding the causes and consequences of social categorization* (Cambridge, Massachusetts, United States: Academic Press), 1–80. doi:10.1016/bs.aesp.2016.10.001
- Kim, Y., and Lim, J. H. (2013). Gendered socialization with an embodied agent: creating a social and affable mathematics learning environment for middle-grade females. *J. Educ. Psychol.* 105, 1164–1174. doi:10.1037/a0031027
- Li, J., Zheng, Z., Wei, X., and Wang, G. (2021). “Faceme: an augmented reality social agent game for facilitating children’s learning about emotional expressions,” in Adjunct Proceedings of the 34th Annual ACM Symposium on User Interface Software and Technology, Virtual Event, USA, October, 2021. doi:10.1145/3474349.3480216
- Likas, A., Vlassis, N., and Verbeek, J. (2003). The global k-means clustering algorithm. *Pattern Recognit.* 36, 451–461. doi:10.1016/S0031-3203(02)00060-2
- Lux, G. (2021). Cool Japan and the Hallyu Wave: the effect of popular culture exports on national image and soft power. [https://digitalcommons.ursinus.edu/eastasia\\_hon/3/#:~:text=Japan%20used%20its%20export%20of,threatening%20image%20of%20national%20culture.](https://digitalcommons.ursinus.edu/eastasia_hon/3/#:~:text=Japan%20used%20its%20export%20of,threatening%20image%20of%20national%20culture.)
- MacLin, O. H., and Malpass, R. S. (2001). Racial categorization of faces: the ambiguous race face effect. *Psychol. Public Policy, Law* 7, 98–118. doi:10.1037/1076-8971.7.1.98
- Maister, L., Slater, M., Sanchez-Vives, M. V., and Tsakiris, M. (2015). Changing bodies changes minds: owning another body affects social cognition. *Trends Cognitive Sci.* 19, 6–12. doi:10.1016/j.tics.2014.11.001
- Maisuwong, W. (2012). The promotion of American culture through Hollywood movies to the world. *Int. J. Eng. Res. Technol. (IJERT)* 1, 1–7.
- Malik, M., Hopp, F. R., and Weber, R. (2022). Representations of racial minorities in popular movies. *Comput. Commun. Res.* 4, 208–253. doi:10.5117/ccr2022.1.006.mali
- Matthews, K., Phelan, J., Jones, N. A., Konya, S., Marks, R., Pratt, B. M., et al. (2017). *National content test race and ethnicity analysis report*. Washington, D.C., United States: U.S. Department of Commerce.
- McIntosh, J., Zajac, H. D., Stefan, A. N., Bergström, J., and Hornbæk, K. (2020). “Iteratively adapting avatars using task-integrated optimisation,” in Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology, Virtual Event, USA, October, 2020, 709–721. doi:10.1145/3379337.3415832
- Meissner, C. A., and Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: a meta-analytic review. *Psychol. Public Policy, Law* 7, 3. doi:10.1037/1076-8971.7.1.3
- Moreno, R., and Flowerday, T. (2006). Students’ choice of animated pedagogical agents in science learning: a test of the similarity-attraction hypothesis on gender and ethnicity. *Contemp. Educ. Psychol.* 31, 186–207. doi:10.1016/j.cedpsych.2005.05.002
- Mori, M., MacDorman, K. F., and Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics Automation Mag.* 19, 98–100. doi:10.1109/MRA.2012.2192811
- Nag, P., and Yalçın, Ö. N. (2020). “Gender stereotypes in virtual agents,” in Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents, Virtual Event Scotland, UK, October, 2020. doi:10.1145/3383652.3423876
- Nicolas, G., Skinner, A. L., and Dickter, C. L. (2019). Other than the sum: hispanic and Middle eastern categorizations of black–white mixed-race faces. *Soc. Psychol. Personality Sci.* 10, 532–541. doi:10.1177/1948550618769591
- Nowak, K. L., and Biocca, F. (2003). The effect of the agency and anthropomorphism on users’ sense of telepresence, copresence, and social presence in virtual environments. *Presence Teleoperators Virtual Environ.* 12, 481–494. doi:10.1162/105474603322761289
- Ogawa, N., Narumi, T., Kuzuoka, H., and Hirose, M. (2020). “Do you feel like passing through walls? effect of self-avatar appearance on facilitating realistic behavior in virtual environments,” in Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu HI, USA, April, 2020, 1–14. doi:10.1145/3313831.3376562
- Palan, S., and Schitter, C. (2018). Prolific.ac—a subject pool for online experiments. *J. Behav. Exp. Finance* 17, 22–27. doi:10.1016/j.jbef.2017.12.004
- Peck, T. C., and Gonzalez-Franco, M. (2021). Avatar embodiment. A standardized questionnaire. *Front. Virtual Real.* 1, 1–12. doi:10.3389/frvir.2020.575943
- Peck, T. C., Good, J. J., and Seitz, K. (2021). Evidence of racial bias using immersive virtual reality: analysis of head and hand motions during shooting decisions. *IEEE Trans. Vis. Comput. Graph.* 27, 2502–2512. doi:10.1109/TVCG.2021.3067767
- Peck, T. C., Seinfeld, S., Aglioti, S. M., and Slater, M. (2013). Putting yourself in the skin of a black avatar reduces implicit racial bias. *Conscious. Cognition* 22, 779–787. doi:10.1016/j.concog.2013.04.016
- Peer, E., Brandimarte, L., Samat, S., and Acquisti, A. (2017). Beyond the Turk: alternative platforms for crowdsourcing behavioral research. *J. Exp. Soc. Psychol.* 70, 153–163. doi:10.1016/j.jesp.2017.01.006
- Pelachaud, C. (2009). Modelling multimodal expression of emotion in a virtual agent. *Philosophical Trans. R. Soc. B Biol. Sci.* 364, 3539–3548. doi:10.1098/rstb.2009.0186
- Piumsomboon, T., Lee, G. A., Hart, J. D., Ens, B., Lindeman, R. W., Thomas, B. H., et al. (2018). “Mini-me: an adaptive avatar for mixed reality remote collaboration,” in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal QC, Canada, April, 2018. doi:10.1145/3173574.3173620
- Rhodes, G., Lie, H. C., Ewing, L., Evangelista, E., and Tanaka, J. W. (2010). Does perceived race affect discrimination and recognition of ambiguous-race faces? A test of the sociocognitive hypothesis. *J. Exp. Psychol. Learn. Mem. Cognition* 36, 217–223. doi:10.1037/a0017680
- Rhodes, G., Locke, V., Ewing, L., and Evangelista, E. (2009). Race coding and the other-race effect in face recognition. *Perception* 38, 232–241. doi:10.1068/p6110
- Salmanowitz, N. (2018). The impact of virtual reality on implicit racial bias and mock legal decisions. *J. Law Biosci.* 5, 174–203. doi:10.1093/jlb/lxy005
- Saneyoshi, A., Okubo, M., Suzuki, H., Oyama, T., and Laeng, B. (2022). The other-race effect in the uncanny valley. *Int. J. Hum. Comput. Stud.* 166, 102871. doi:10.1016/j.ijhcs.2022.102871
- Scarpina, F., Serino, S., Keizer, A., Chirico, A., Scacchi, M., Castelnovo, G., et al. (2019). The effect of a virtual-reality full-body illusion on body representation in obesity. *J. Clin. Med.* 8. doi:10.3390/jcm8091330
- Setoh, P., Lee, K. J., Zhang, L., Qian, M. K., Quinn, P. C., Heyman, G. D., et al. (2019). Racial categorization predicts implicit racial bias in preschool children. *Child. Dev.* 90, 162–179. doi:10.1111/cdev.12851

- Shapiro, A., Feng, A., Wang, R., Li, H., Bolas, M., Medioni, G., et al. (2014). Rapid avatar capture and simulation using commodity depth sensors. *Comput. Animat. Virtual Worlds* 25, 201–211. doi:10.1002/cav.1579
- Sheskin, D. J. (2011). *Handbook of parametric and nonparametric statistical procedures*. 5th edn. New York, NY, USA: Chapman and Hall/CRC. doi:10.1201/9780429186196
- Sra, M., and Schmandt, C. (2015). "Metaspace: full-body tracking for immersive multiperson virtual reality," in Adjunct Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology, Daegu Kyungpook, Republic of Korea, November, 2015. doi:10.1145/2815585.2817802
- Taylor, V. J., Valladares, J. J., Siepser, C., and Yantis, C. (2020). Interracial contact in virtual reality: best practices. *Policy Insights Behav. Brain Sci.* 7, 132–140. doi:10.1177/2372732220943638
- Thanikachalam, S., Sanchez, N., and Maddy, A. J. (2019). Long hair throughout the ages. *Dermatology* 235, 260–262. doi:10.1159/000497156
- Volonte, M., Ofek, E., Jakubzak, K., Bruner, S., and Gonzalez-Franco, M. (2022). "Headbox: a facial blendshape animation toolkit for the microsoft rocketbox library," in 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Christchurch, New Zealand, March, 2022, 39–42. doi:10.1109/VRW55335.2022.00015
- Waltemate, T., Gall, D., Roth, D., Botsch, M., and Latoschik, M. E. (2018). The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE Trans. Vis. Comput. Graph.* 24, 1643–1652. doi:10.1109/TVCG.2018.2794629
- Yee, N., and Bailenson, J. (2007). The Proteus effect: the effect of transformed self-representation on behavior. *Hum. Commun. Res.* 33, 271–290. doi:10.1111/j.1468-2958.2007.00299.x
- Zipp, S. A., and Craig, S. D. (2019). The impact of a user's biases on interactions with virtual humans and learning during virtual emergency management training. *Educ. Technol. Res. Dev.* 67, 1385–1404. doi:10.1007/s11423-019-09647-6